Cyber Security Issues of HEP and NP Grids

Author/Contact: Bob Cowles, rdc@slac.stanford.edu
Collaboratory: PPDG
6 July 2004
For National Collaboratories Meeting, 10-12, August 2004

A broad community of groups is working together for safe grid computing in the HENP environment; however, there remains a significant cyber security impact on researchers, application developers, research organizations and sites. For each of these groups, we discuss what are currently seen as the major changes necessary to successfully deploy secure grid services in an open research environment.

Researchers

Some form of secure, long-term identity token repository will be required. For now, many sites are using X.509 credentials stored in file systems, but we see a need for acceptance of a variety of authentication methods including smart cards and One-Time Password (OTP) tokens. The latter forms of authentication would be performed by one site but accepted by other sites willing to trust another's credentials for access to a limited set of resources.

Registration with a research organization provides the link between the identity token and the ability to use resources of the organization on the grid. As part of the registration process, sufficient information will need to be collected so that people responsible for security feel confident in their ability to get in contact with the researcher "reasonably quickly" in case of a problem.

Ultimately, sites are in control of the resources they allow access to, so researchers must set up their requests to accurately specify minimum resource requirements for a task, adapt to changes in resources that may be available at a given time and access those resources only through high-level interfaces specifically designed for grid applications. Not following these guidelines will guarantee tasks fail for what appear to be strange and mysterious reasons. The implementation and deployment of resource authorization is new to everyone (researchers, sites, research organizations) so it's best to keep things simple at the start.

Application Developers

Since sites are in control of access to their resources, there are a variety of policies with respect to application access to network, storage and computational resources. Application developers must be aware that an absolute requirement for certain kinds of resource access (e. g. outgoing IP connectivity) may severely limit the number of sites willing to run such an application. Negotiation at the design stage on how adequate security can be provided (e. g. use of proxy services) can dramatically increase the number of sites willing to run applications.

To resolve potential security problems and track down distributed processing errors, applications need to log "significant events" that occur in the course of their processing. Logging should be done in a standardized form to a standard location so that logs from a variety of distributed applications and services can be reasonably analyzed to develop a trace of a task through a grid. Giving the researcher appropriate troubleshooting information from the ensemble of system and application logs while not providing too much information to attackers will be difficult.

Due to the distributed nature of processing on the grid, applications and services must be designed to be robust with respect to the temporary unavailability of a required resource (either cause by a network glitch or because the service had to be shutdown and restarted). The reason this is a security issue is that services must be able to be restarted to apply security patches and to upgrade version for security reasons.

Software for applications and services should be developed with secure programming practices. Included, of course, is checking all input for possible errors and verifying the program reacts robustly when given unexpected input or changes in its environment. A new feature of this environment is the need to periodically test if the authentication or authorization has been revoked. While this can be difficult to implement, it is a necessary control to prevent misuse of credentials or resources from going on forever.

Research Organizations

Sites are going to have to depend on the research organizations, much more than in the past, to maintain an accurate description of who is allowed to use the sites' resources and at what level of privilege (e. g. what role is being assumed). Critical databases containing this information will be continually accessed to obtain authorization information so procedures must be in place to maintain the database hardware, software and data in a secure, reliable and auditable fashion.

For any task submitted by a researcher, the computation and results must be traceable back to the individual who made the request. Therefore, any mapping of the original identity into a different individual or group identity must be logged and must have a reverse mapping so that any resulting problems can be traced back to the source.

All resource and privilege or role decisions must be logged and hooks provided for intrusion detection. Again, standardized formats must be used to allow reasonable tracing of a task request through a grid and the variety of resources and services it makes use of.

Sites

The increased power of grid comes at the price of increased interdependency between sites. Sufficient control over use of resource must be exercised to detect unauthorized or anomalous behavior. Rapid reporting of any such potential incidents to other site managers and to the research organization is an absolute requirement for maintaining a safe computing environment.

Rapid application of security patches or otherwise mitigating any security vulnerabilities is necessary to maintain the integrity of the grid infrastructure. Compromised machines or services must be isolated immediately when discovered to help limit the impact or any intrusion. Compromised or suspected compromised identities must be reported immediately to the research organization for action and to other sites for actions according to local site policies.